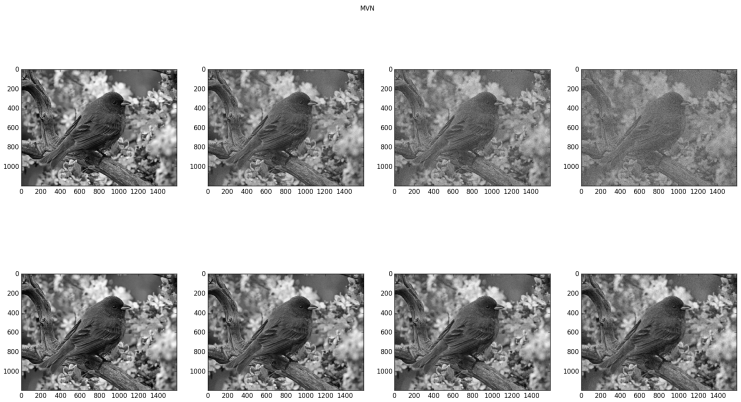
Advanced Practical Machine Learning

Keren Meron 200039626

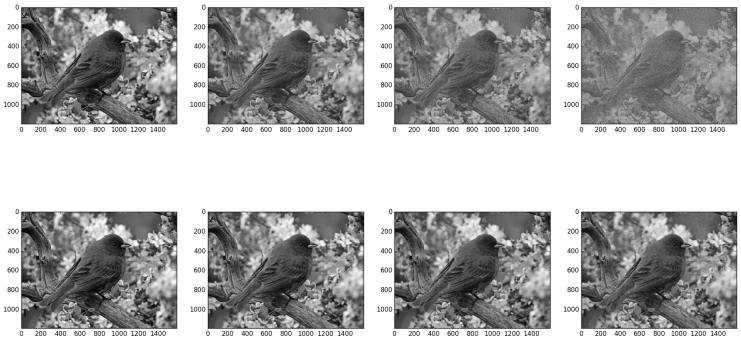
November 30, 2017

Results:

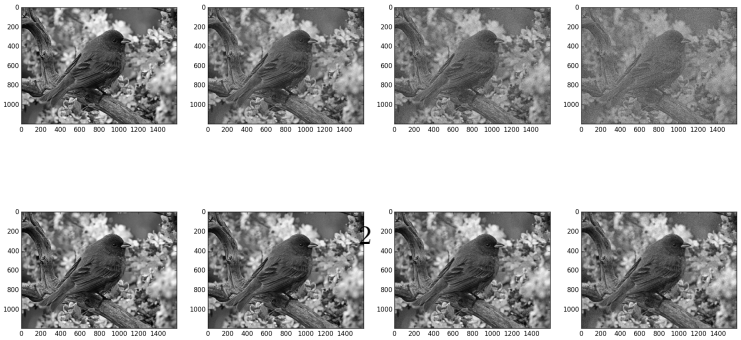
MVN



GSM



ICA



Discussion

The best model is the GSM model.

The ICA model does not give as good results because by looking at each dimension separetly, it does not take into consideration the surroundings of each pixel. In an image, it is very wrong to assume that each pixel 'dimension' holds by itself and is not correlated to the other, and in particular its neighboring, pixels in the image. Therefore, we lose a lot of information using the ICA model, which is why it gives more poor results than the GSM model.

The MVN model does not work as well as the GSM because it is a much simpler model. GSM can be looked at as an extended model of the MVN, where instead of assuming the data distributes as some guassian distribution, we allow to look at several k gaussian, and denoise the image according to a weighted expectation of k learnt models. We know that the MVN assumption is not percise, and that it is inflicted in the result.

Ampirically, we can also see in the above results that the GSM denoising has the "nicest" denoised image.

Run Time Analysis

The average times it took to run the three algorithms (total learn + denoise), performed on the same hardware, were:

- MVN: 19.7133 seconds
- GSM: 106.6427 seconds
- ICA: 269.2611 seconds

The average times it took to run the three algorithms (only learn), performed on the same hardware, were:

- MVN: 0.00977 seconds
- GSM: 0.2072 seconds
- ICA: 1.3853 seconds

Obviously we see that MVN is by far the fastest, as it has straight computing and no k dependance. ICA takes much longer than the others, as it performs the learning for each dimension d separately, thus taking much more time.

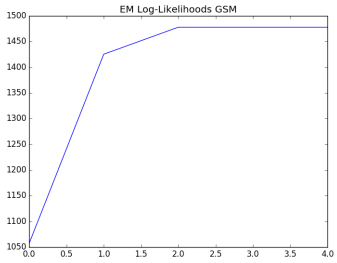
Log Likelihoods

The final LLEs for each algorithm on clean test data (after adding a (big) constant):

- MVN: 728720638.775
- GSM: -349114911.7
- ICA: 63055000683.1

We get that test data is most similar to the ICA model, afterwards to the MVN and finally to GSM.

The convergence of the Expectation Maximization algorithm can be seen in the following graph of the log likelihoods calculated at each iteration of the algorithm:



We can see that the algorithm quickly converges to a solution.

MSE

noise	MVN	GSM	ICA	WINNER
noisy 0.01	0.000100119930519	9.98036293627e-05	9.99334192183e-05	GSM
denoised 0.01	0.00014637084818	0.000145938159409	0.000113470138886	/
noisy 0.05	0.00250569008002	0.00249752145235	0.00250383047165	GSM
denoised 0.05	0.000710623399393	0.000704826540514	0.000641611501717	/
noisy 0.1	0.00999622606503	0.00998818823029	0.0100029625999	GSM
denoised 0.1	0.00117976125485	0.00117851553432	0.00114221101033	/
noisy 0.2	0.0400195047597	0.0399596279633	0.040015448522	GSM
denoised 0.2	0.00185356256209	0.00185325399192	0.00190355634265	/